

Estimating Sparse Signals Using Integrated Wideband Dictionaries

Maksim Butsenko*, Johan Swärd†, and Andreas Jakobsson†

Abstract—In this paper, we introduce a wideband dictionary framework for estimating sparse signals. By formulating integrated dictionary elements spanning bands of the considered parameter space, one may efficiently find and discard large parts of the parameter space not active in the signal. After each iteration, the zero-valued parts of the dictionary may be discarded to allow a refined dictionary to be formed around the active elements, resulting in a zoomed dictionary to be used in the following iterations. Implementing this scheme allows for more accurate estimates, at a much lower computational cost, as compared to directly forming a larger dictionary spanning the whole parameter space or performing a zooming procedure using standard dictionary elements. Different from traditional dictionaries, the wideband dictionary allows for the use of dictionaries with fewer elements than the number of available samples without loss of resolution. The technique may be used on both one- and multi-dimensional signals, and may be exploited to refine several traditional sparse estimators, here illustrated with the LASSO and the SPICE estimators. Numerical examples illustrate the improved performance.

I. INTRODUCTION

A wide range of common applications yield signals that may be well approximated using a sparse reconstruction framework, and the area has as a result attracted notable interest in the recent literature (see, e.g., [1]–[3] and the references therein). Much of this work has focused on formulating convex algorithms that exploit different sparsity inducing penalties, thereby encouraging solutions that are

well represented using only a few elements from some (typically known) dictionary matrix, D . If the dictionary is appropriately chosen, even very limited measurements can be shown to allow for an accurate signal reconstruction [4], [5]. Recently, increasing attention has been given to signals that are best represented using a continuous parameter space. In such cases, the discretization of the parameter space that is typically used to approximate the true parameters will not represent the noise-free signal exactly, resulting in solutions that are less sparse than desired. This problem has been examined in, e.g., [6]–[8], wherein discretization recommendations and new bounds of the reconstruction guarantees were presented, taking possible grid mismatches into consideration. Typically, this results in the use of large and over-complete dictionaries, which, although quite efficient, often violate the assumptions required to allow for a perfect recovery guarantee.

As an alternative, one may formulate the reconstruction problem using a continuous dictionary, such as in, e.g., [9]–[11]. This kind of formulations typically use an atomic norm penalty, as introduced in [12], which allows for a way to determine the most suitable convex penalty to recover the signal, even over a continuous parameter space. These solutions often offer an accurate signal reconstruction, but also require the solving of large and computationally rather cumbersome optimization problems, thereby limiting the size of the considered problems.

In this work, we examine an alternative way of approaching the problem, proposing the use of wideband dictionary elements, such that the dictionary is formed over B subsets of the continuous parameter space. In the estimation procedure, the activated subsets are retained and refined, whereas non-activated sets are discarded from the further optimization. This screening procedure may be broken down into two steps. The first step is to remove the parts of the parameter space not active in the

This work was supported in part by the Swedish Research Council and Crafoord's foundation and in part by the ERA-chair project COEL grant (No 668995) financed by European Unions Horizon 2020 research and innovation program, the Institutional Research Project IUT19-11 financed by the Estonian Research Council, and the project B38, financed by the Tallinn University of Technology. Parts of the material herein have been presented at the 2017 ICASSP conference.

*Thomas Johann Seebeck Dept. of Electronics, Tallinn University of Technology, Ehitajate tee 5, 19086, Tallinn, Estonia, email: maksim.butsenko@ttu.ee.

†Dept. of Mathematical Statistics, Lund University, P.O. Box 118, SE-221 00 Lund, Sweden, email: {js, aj}@maths.lth.se.

signal, whereafter, in the second step, a smaller dictionary is formed covering only the parts of the parameter space that were active in the first step. This smaller dictionary may then again be expanded with candidates close to the activated elements, thereby yielding a zoomed dictionary in these regions. The process may then be repeated to further refine the estimates as desired. Without loss of generality, the proposed principle is here illustrated on the problem of estimating the frequencies of K complex-valued M -dimensional sinusoid corrupted by white circularly symmetric Gaussian noise. The one-dimensional case of this is a classical estimation problem, originally expressed using a sparse reconstruction framework in [13], and having since attracting notable attention (see, e.g., [14]–[17]). Here, using the classical formulation, the resulting sinusoidal dictionary will allow for a K -sparse representation of frequencies on the grid, whereas the grid mismatch of any off-grid components will typically yield solutions with more than K components. Extending the dictionary to use a finely spaced dictionary, as suggested in, e.g., [8], will yield the desired solution, although at the cost of an increased complexity. In this work, we instead proceed to divide the spectrum into B (continuous) frequency bands, each band possibly containing multiple spectral lines. This allows for an initial coarse estimation of the signal frequencies, without the risk of missing any off-grid components. Due to the iterative refining of the dictionary, closely spaced components are successfully separated as the dictionary is refined; as the wideband elements span the full band, no power is off-grid, avoiding the problem of a non-sparse solution due to dictionary mismatch.

Other screening methods that decrease the dictionary size have been proposed. For instance, in [18]–[23], methods for finding the elements in the dictionary that corresponds to zero-valued elements in the sparse vector were proposed. Based on the inner product between the large dictionary and the signal, a rule was formed for deeming whether or not a dictionary element was present in the signal or not. Although these methods show a substantial decrease in computational complexity, one still has to form the inner product between the likely large dictionary and the signal. To alleviate this, one may instead use the here proposed wideband dictionary elements, thereby discarding large parts of the pa-

rameter space. Since the wideband dictionary is magnitudes smaller than the full dictionary required to achieve the reconstruction, the computational complexity is significantly reduced.

The proposed principle is not limited to methods that use discretization of the parameter space; it may also be used when solving the reconstruction problem using gridless methods, such as the methods in [9]–[11]. It has been shown that if the reconstruction problem allows for any prior knowledge about the location of the frequencies, e.g., the frequencies are located within a certain region of the spectrum, one may use this information to improve the estimates [24]. The proposed method may also be used to attain such prior information, and thus improving the overall estimates as a result.

To illustrate the performance of the proposed dictionary, we make use of two different sinusoidal estimators, namely the LASSO [25] and the SPICE estimators [26], [27]; the first finding the estimate by solving a penalized regression problem, whereas the latter instead solves a covariance fitting problem.

The remainder of this paper is organized as follows: in the next section, the problem of estimating an M -dimensional sinusoidal signal is introduced, followed, in Section III, by the introduction of the proposed wideband dictionary. In Section IV, a discussion about the computational complexity reduction allowed by the proposed wideband dictionary is given, and, in Section V, the performance of the proposed wideband dictionary is illustrated by numerical examples. Finally, in Section VI, we conclude on our work.

II. PROBLEM STATEMENT

To illustrate the wideband dictionary framework consider the problem of estimating the K frequencies $f_k^{(m)}$, for $k = 1, \dots, K$ and $m = 1, \dots, M$, of an M -dimensional signal y_{n_1, \dots, n_M} , with

$$y_{n_1, \dots, n_M} = \sum_{k=1}^K \beta_k e^{2i\pi f_k^{(1)} t_{n_1}^{(1)} + \dots + 2i\pi f_k^{(M)} t_{n_M}^{(M)}} + \epsilon_{n_1, \dots, n_M} \quad (1)$$

for $n_m = 1, \dots, N_m$, and where K denotes the (unknown) number of sinusoids in the signal. Furthermore, let β_k and $f_k^{(m)}$ denote the complex amplitude and frequency of the k th frequency and m th dimension, respectively, $t_{n_m}^{(m)}$ the n_m th sample time in the m th dimension, and $\epsilon_{n_1, \dots, n_M}$ an additive noise

observed at time t_{n_1}, \dots, t_{n_M} . The signal model in (1) may be equivalently described by an M -dimensional (M -D) tensor

$$\mathbf{y} = \sum_{k=1}^K \beta_k \tilde{\mathbf{d}}_{(k)}^{(1)} \circ \tilde{\mathbf{d}}_{(k)}^{(2)} \cdots \circ \tilde{\mathbf{d}}_{(k)}^{(M)} + \boldsymbol{\varepsilon} \quad (2)$$

where \circ denotes the outer product, and

$$\tilde{\mathbf{d}}_{(k)}^{(m)} = \begin{bmatrix} e^{2i\pi f_k^{(m)} t_1^{(m)}} & \dots & e^{2i\pi f_k^{(m)} t_{N_m}^{(m)}} \end{bmatrix}^T \quad (3)$$

To determine the parameters of the model in (1) or (2), as well as the model order, we proceed by creating a dictionary containing a set of signal candidates, each representing a sinusoid with a unique frequency. By measuring the distance between the signal candidates and the measured signal, and by promoting a sparse solution, one may find a small set of candidates that best approximates the signal. To this end, we form dictionary elements on the form

$$\mathbf{d}_{(k)}^{(m)} = \begin{bmatrix} e^{2i\pi f_p^{(m)} t_1^{(m)}} & \dots & e^{2i\pi f_p^{(m)} t_{N_m}^{(m)}} \end{bmatrix}^T \quad (4)$$

for $p = 1, \dots, P_M$, where $P_M \gg K$ denotes the number of candidates in dimension m . Here, the dictionary is assumed to be fine enough so that the unknown sinusoidal component will (reasonably well) coincide with K dictionary elements. Often, it is more convenient to work with a vectorized version of the tensor. Let $\mathbf{y} = \text{vec}(\mathbf{y})$, where $\text{vec}(\cdot)$ stacks the tensor into a vector. One may then re-write (2) as

$$\mathbf{y} = (\mathbf{D}^{(M)} \otimes \mathbf{D}^{(M-1)} \otimes \dots \otimes \mathbf{D}^{(1)}) \boldsymbol{\beta} \quad (5)$$

where \otimes denotes the Kronecker product, suggesting that one may find both the unknown parameters and the model order by forming the LASSO problem (see, e.g., [13], [25])

$$\min_{\boldsymbol{\beta}} \|\mathbf{y} - \mathbf{D}\boldsymbol{\beta}\|_2^2 + \lambda \|\boldsymbol{\beta}\|_1 \quad (6)$$

where $\mathbf{D} = (\mathbf{D}^{(M)} \otimes \mathbf{D}^{(M-1)} \otimes \dots \otimes \mathbf{D}^{(1)})$ and $\|\cdot\|_q$ denotes the q -norm. A visual representation of such dictionary is shown in Figure 1 for the 2-D case. The penalty on the 1-norm of $\boldsymbol{\beta}$ will ensure that the found solution will be sparse, with λ denoting a user parameter governing the desired sparsity level of the solution. The frequencies, as well as their order, are then found as the non-zero elements in $\boldsymbol{\beta}$.

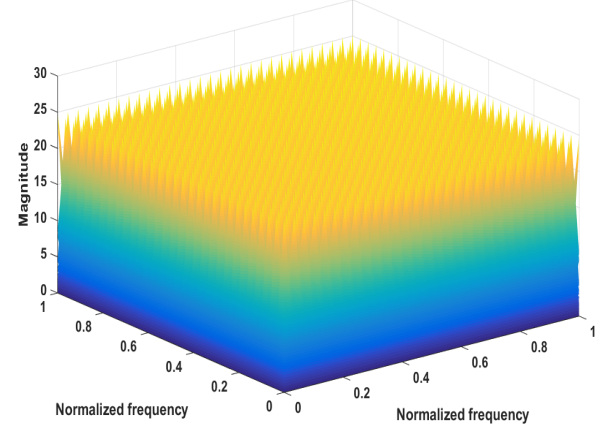


Fig. 1: Fine-grid dictionary for two-dimensional signal estimation with $N_1 = 30$, $N_2 = 30$, and $P = 60$ elements per dimension.

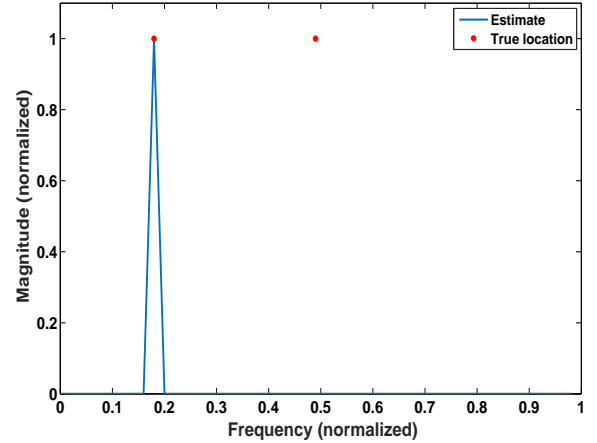


Fig. 2: The inner-product of a dictionary containing $P = 50$ (narrowband) candidate frequency elements and the noise-free signal, with $N = 100$.

As shown in [8], the number of dictionary elements, P , typically has to be large to allow for an accurate determination of the correct parameters. This means that for multi-dimensional signals, the dictionary quickly becomes inhibitory large. Thus, it is often not feasible in practice to directly compute the solution of (6) using a dictionary constructed from such finely space candidates. As an alternative, one may use a zooming procedure, where one first employs an initial coarse dictionary, \mathbf{D}_1 , to determine the parameter regions of interest, and then employ a fine dictionary, \mathbf{D}_2 , centered around the initially found candidates (see, e.g., [28], [29] for similar approaches). This allows for a com-

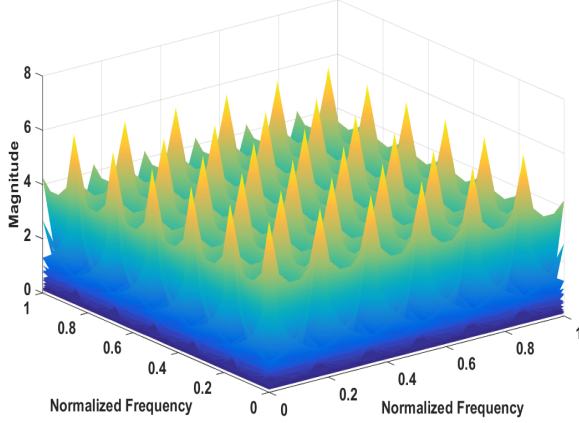


Fig. 3: Wideband dictionary with integrated sinusoids elements formed with $N_1 = 30$, $N_2 = 30$, and $B = 6$ bands per dimension.

putationally efficient solution of the optimization problem in (6), but suffers from the problem of possibly missing off-grid components far from the initial coarse frequency grid. This is illustrated in Figure 2 for a 1-D signal, where the inner-product between the dictionary and the signal is depicted together with the location of the true peaks. In this noise-free example, we used $N = 100$ samples and $P = 50$ dictionary elements, with one of the frequencies being situated in between two adjacent grid points in the dictionary. As seen in the figure, the coarse initial estimate fails to detect the presence of the second signal component, which is thereby discarded as a possibility in the following refined estimate. Increasing the number of candidate frequencies will result in the side-lobes of the dictionary elements decreasing the gap between the frequency grid points, making the inner-product between the dictionary and the signal larger for components that lie in between two candidate frequencies. However, doing so will increase the computational complexity correspondingly, begging the question if one may retain a low number of candidate frequencies, while still reducing the likelihood of missing any off-grid components. This is the problem we shall examine in the following.

III. INTEGRATED WIDEBAND DICTIONARIES

We note that the above problem results from the dictionary being formed over a set of single-component candidates, thereby increasing the risk

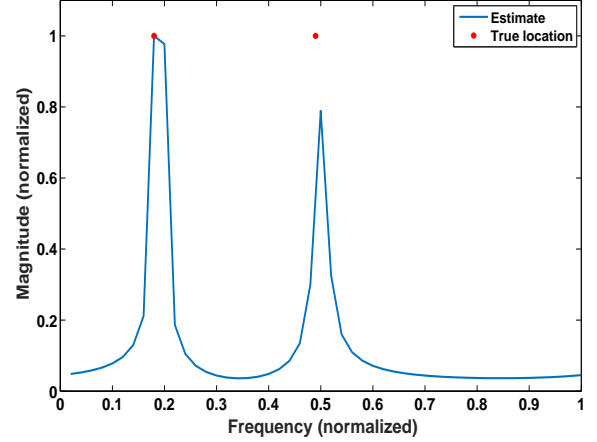


Fig. 4: The inner-product of a dictionary containing $B = 50$ (wideband) candidate frequency elements and the noise-free signal, with $N = 100$.

of neglecting the off-grid components. In order to avoid this, we here propose a wideband dictionary framework, such that each of the dictionary elements is instead formed over a range of such single-component candidates. This is done by letting the dictionary elements be formed over an integrated range of the parameter(s) of interest, in this case being the frequencies of the candidate sinusoids. For a multi-dimensional sinusoidal dictionary, the resulting B integrated wideband elements should thus be formed as

$$a_{b(1), \dots, b(M)}(t^{(1)}, \dots, t^{(M)}) = \int_{f_b(1)}^{f_{b(1)+1}} \dots \int_{f_b(M)}^{f_{b(M)+1}} e^{2i\pi(f^{(1)}t^{(1)} + \dots + f^{(M)}t^{(M)})} df^{(1)} \dots df^{(M)} \quad (7)$$

for $t^{(m)} = 1, \dots, N_m$ for all $m = 1, \dots, M$, where $f_b^{(m)}$ and $f_{b+1}^{(m)}$ are the two frequencies bounding the frequency band, for $b = 1, \dots, B$, for the m th dimension. The resulting elements are then gathered into the dictionary, \mathbf{B} , where each column contains a specific wideband of the M -D parameter space for all time samples, where each element is formed as the solution from (7), such that, in this case,

$$\prod_{m=1}^M \frac{e^{2i\pi f_{b(m)+1} t^{(m)}} - e^{2i\pi f_{b(m)} t^{(m)}}}{2i\pi t^{(m)}} \quad (8)$$

Note that (8) corresponds to the M -D inverse

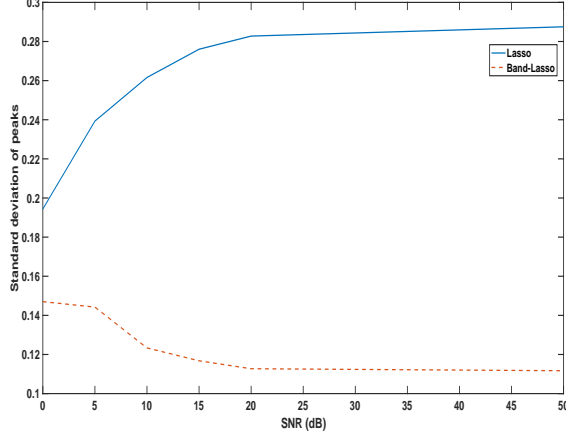


Fig. 5: The standard deviation of the peaks as a function of SNR.

Fourier transform of 1, i.e., it is the M -D inverse Fourier transform of an M -D section in the frequency domain with unit amplitude. For the 1-D case, this simplifies to

$$\begin{cases} 1, & \text{for } f_a \leq f \leq f_b \\ 0 \end{cases} \xrightarrow{\mathcal{F}^{-1}} \begin{cases} \frac{e^{2i\pi f_b t} - e^{2i\pi f_a t}}{2i\pi t} \\ 0 \end{cases} \quad (9)$$

In Figure 3, we show a visual representation of the resulting wideband dictionary for $M = 2$ dimensions. The inner-product between the proposed dictionary, \mathbf{B} , and the earlier 1-D signal is shown in Figure 4, using the same number of dictionary elements as in Figure 1, clearly indicating that the proposed dictionary is able to locate the off-grid frequency. This is due the wideband nature of the proposed dictionary, which thus has less power concentrated at the grid points, but covers a wider range of frequencies, not reducing to zero, or close to zero, anywhere within the band (as is the case for the narrowband dictionary elements). As a result, using the wideband dictionary elements, it is possible to use a smaller dictionary, thereby reducing the computational complexity, without increasing the risk of missing components in the signal. To further show this, 1000 Monte-Carlo simulations were conducted for each considered SNR-level. In each simulation, we considered a signal containing two sinusoids, where the frequencies were randomly selected on $(0, 1]$ with a spacing of at least $2/N$, with $N = 100$ denoting the signal length. The sinusoids had the magnitudes 4 and 5, with a randomly selected phase

between $(0, 2\pi]$. Two dictionaries were given, one containing ordinary sinusoids and one containing the proposed wideband components, both containing $P = B = 50$ elements. For each dictionary, the inner-products with the signal were computed, where the amplitudes were normalized so that the largest estimated peak had unit magnitude. Figure 5 shows the variance of the smallest peak for different SNR-levels. As is clear from the figure, the variance of the peaks are much lower for the banded case. The reason why the sinusoidal dictionary results in a larger variance is due to the fact that the main lobe is much thinner in this case than in the banded counterpart. This means that when the sinusoids happen to have frequencies that do not overlap with the main lobe of the dictionary, the power in the inner-product will be small. This will not only make such components harder to detect, but will also make it more difficult to determine a suitable regularizing hyperparameter, λ .

When P decreases below N , the gaps between the frequency candidates in the single-component dictionary become so large that if one of the sinusoids in the signal has its frequency values between two adjacent grid points, the likelihood that this sinusoid lie in the null-space of the dictionary increases. This problem is avoided with the wideband dictionary as it is more likely to eliminate any gaps.

This property is depicted in Figure 6, where the success rate of finding the true support is displayed as a function of the number of samples, N , and the number of bands in the dictionary, B , for different number of sinusoids in the signal, K . The estimation was done for a noise-free signal by solving (6), using wideband dictionaries and letting

$$\lambda = 0.3 \max_{i=1, \dots, B} |\mathbf{d}_i \mathbf{y}| \quad (10)$$

where \mathbf{d}_i denotes the i th column of \mathbf{D} . In the top left figure, the signal contains three sinusoids, and it is clearly the case that the banded dictionary is unable to retrieve the true support for all setting of N and M/N , except for the case when $N = 30$ and $M/N < 7$. In the top right and bottom figures, where $K = 7$ and $K = 11$, respectively, it is shown that when the number of sinusoids in the signal increases, a larger number of samples is needed to allow for a successful reconstruction, which is reasonable, as one needs more information to be able to correctly estimate more parameters.

However, the banded dictionary is able to retrieve the true support as long as the number of samples is big enough and the ratio M/N is not too small. It is further clear from the figures, that the banded dictionary actually retrieves the true support even though $M < N$. Examining the part of the signal that is unexplained by the support, one may note that no bands outside the true support were activated. Thus, only the bands that either were part of the true support, or that were adjacent to a band included in the true support, were activated. The reason why some of the adjacent bands were activated is that when the true frequency is very close to the left (or right) limit of the band, it will also activate the adjacent bands.

The proposed approach is not the only way to form a wideband dictionary. For example, one could populate the dictionary using discrete prolate spheroid sequences (DPSS) [30]. For an integer Q and with real-valued $0 < W < \frac{1}{2}$, the DPSS are a set of Q discrete-time sequences for which the amplitude spectrum is band-limited. The most interesting property of the DPSS for our discussion is the fact that the energy spectrum of the dictionary elements are highly concentrated in the range $[-W, W]$, suggesting that the DPSS could be a suitable basis for the candidates in a wideband dictionary, where the candidates are formed such that each covers a $1/B$ -th part of the spectrum. In the numerical section below, we examine how the use of DPSS candidates compare to the integrated wideband candidates in (8).

IV. COMPLEXITY ANALYSIS

To illustrate the computational benefits of using the wideband dictionary as compared to forming the full dictionary, we proceed with our example of determining K M -D sinusoids by solving (6) using the popular ADMM algorithm [31]. In order to do so, we first transform the problem into a vector form reminiscent to (5), and split the variable β into two variables, here denoted \mathbf{x} and \mathbf{z} , after which the optimization problem may be reformulated as

$$\min_{\mathbf{x}, \mathbf{z}} \frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda \|\mathbf{z}\|_1 \quad \text{subj. to} \quad \mathbf{x} = \mathbf{z} \quad (11)$$

having the (scaled) augmented Lagrangian

$$\frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda \|\mathbf{z}\|_1 + \frac{\rho}{2} \|\mathbf{x} - \mathbf{z} + \mathbf{u}\|_2^2 \quad (12)$$

where \mathbf{u} is the scaled dual variable and ρ is the step length (see [31] for a detailed discussion on the ADMM). The minimization is thus formed by iteratively solving (12) for \mathbf{x} and \mathbf{z} , as well as updating the scaled dual variable \mathbf{u} . This is done by finding the (sub-)gradient for \mathbf{x} and \mathbf{z} of the augmented Lagrangian, and setting it to zero, fixing the other variables to their latest values. The steps for the j th iteration are thus

$$\mathbf{x}^{(j+1)} = (\mathbf{A}^H \mathbf{A} + \rho \mathbf{I})^{-1} (\mathbf{A}^H \mathbf{y} + \mathbf{z}^{(j)} - \mathbf{u}^{(j)}) \quad (13)$$

$$\mathbf{z}^{(j+1)} = S(\mathbf{x}^{(j+1)} + \mathbf{u}^{(j)}, \lambda/\rho) \quad (14)$$

$$\mathbf{u}^{(j+1)} = \mathbf{u}^{(j)} + \mathbf{x}^{(j+1)} - \mathbf{z}^{(j+1)} \quad (15)$$

where $(\cdot)^H$ denotes the Hermitian transpose, $(\cdot)^{(j)}$ the j th iteration, and $S(\mathbf{v}, \kappa)$ is the soft threshold operator, defined as

$$S(\mathbf{v}, \kappa) = \frac{\max(|\mathbf{v}| - \kappa, 0)}{\max(|\mathbf{v}| - \kappa, 0) + \kappa} \odot \mathbf{v} \quad (16)$$

where \odot denotes the element-wise multiplication for any vector \mathbf{v} and scalar κ .

The computationally most demanding part of the resulting ADMM implementation is to form the inverse in (13) and to calculate $\mathbf{A}^H \mathbf{y}$. These steps are often done by QR factorizing the inverse in (13) prior to the iteration, so that this part is only calculated once. After this, the QR factors are used when forming the inner product. To give a simple example on the difference between the two types of dictionaries, we exclude any further computational speed-ups and show the difference on brute force computations of the above ADMM. This is done to give an idea on the effect $P < N$ has on the computational complexity. The total computational cost for the step in (13) depends on the size of the matrix \mathbf{A} . Let $N = \prod_{m=1}^M N_m$ and $P = \prod_{m=1}^M P_m$, then \mathbf{A} is a $N \times P$ matrix. If $P < N$, computing the inverse will cost approximately P^3 operations, plus an additional $P^2 N$ operations to form the Gram matrix $\mathbf{A}^H \mathbf{A}$. Furthermore, to compute $\mathbf{A}^H \mathbf{y}$ requires PN operations, and the final step to compute \mathbf{x} costs P^2 operations. If instead $P > N$, one may make use of the Woodbury matrix identity [32], allowing the inverse to be formed using $N^3 + 3PN^2$ operations, whereafter one has to compute $\mathbf{A}^H \mathbf{y}$ and the final matrix-vector multiplication, together costing $PN + P^2$ operations. In total, the x-step will

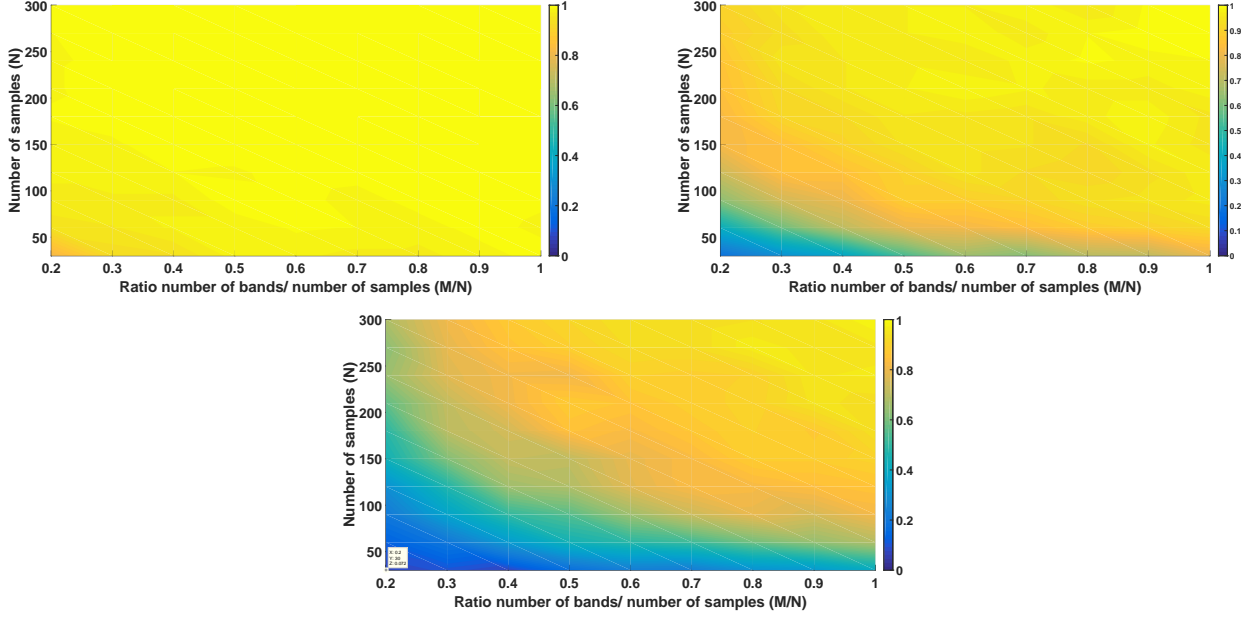


Fig. 6: The success rate of finding the true support as a function of the number of samples (y-axis) and the ratio between the number of bands in the dictionary and the number of samples (x-axis), for different values of K . Top left corner, $K = 3$, top right corner, $K = 7$, and bottom, $K = 13$.

have the cost of roughly $P^3 + (N + 1)P^2 + NP$, if $P < N$, or $N^3 + 3PN^2 + PN + P^2$, if $N < P$.

Since using the banded dictionary allows for a smaller dictionary, one may calculate the computational benefit of using the integrated dictionary as compared to just using an ordinary dictionary with large P . Consider using only a single-stage narrowband dictionary, \mathbf{D}_1 , with $P > N$ dictionary elements. This requires $C_1 = N^3 + 3PN^2 + P^2 + PN$ operations if using the above ADMM solution, with the dictionary \mathbf{D}_1 in the place of \mathbf{A} in (13)-(15). If, on the other hand, one uses a multiple-stage wideband dictionary with N dictionary elements in the initial coarse dictionary, \mathbf{B}_1 (which is more than required, but simplifies the calculations), the cost of forming the first stage (coarse) minimization is $C_2 = 2(N^3 + N^2)$. By taking the difference, i.e., forming

$$R = C_1 - C_2 = N^3 + 3PN^2 + P^2 + PN - 2(N^3 + N^2)$$

one obtains the available computational resources, R , that are left for the dictionaries of the zoomed-in stages, without increasing the overall computational cost above that of the narrowband dictionary solution. Let \mathbf{B}_z denote the zoomed-in dictionary with ηN number of bands, where $0 < \eta < 1$ denotes the

ratio between the number of available bands in the dictionary and the number of samples. Then, one may deduce the grid size for each \mathbf{B}_z that is allowed without increasing the overall computational complexity as compared to using the narrowband dictionary by solving

$$R = KI_z ((\eta N)^3 + (N + 1)(\eta N)^2 + \eta N^2)$$

where I_z denotes the number of zooming steps and K the number of sinusoids in the signal. To illustrate the resulting difference, consider the following settings: $P = 1000$, $N = 100$, $K = 5$, and $\eta = 2/3$. To only use half the resources that are needed to solve the full narrowband problem, one may, using the wideband dictionary, use 4 stages of zooming, resulting in a grid spacing of roughly 10^{-9} , as compared to 10^{-3} for the narrowband dictionary. One may of course also use a zooming procedure when using the narrowband dictionaries, although this would increase the risk of missing any off-grid component. This means that the smallest number of dictionary elements, for the narrowband dictionary to avoid missing any off-grid components, is $P = N$, and thus the wideband dictionary would need only at most η^2 of the computational resources needed for the ordinary dictionary, at each zooming stage.

It is worth stressing that the wideband dictionary framework introduced here is not limited to the LASSO-style minimizations such as the one examined in (6). There are many other popular methods that could be implemented using this approach. As an example of how the wideband dictionary can be applied for other typical sparse estimation algorithms, consider the SPICE algorithm [14], [27], formed as the solution to

$$\underset{\mathbf{p} \geq 0}{\text{minimize}} \quad \mathbf{y}^* \mathbf{R}^{-1} \mathbf{y} + \|\mathbf{p}\|_1 + \|\boldsymbol{\sigma}\|_1 \quad (17)$$

where

$$\mathbf{R}(\mathbf{p}) = \mathbf{A} \mathbf{p} \mathbf{A}^* \quad (18)$$

$$\mathbf{A} = \begin{bmatrix} \mathbf{B} & \mathbf{I} \end{bmatrix} \quad (19)$$

$$\tilde{\mathbf{p}} = \begin{bmatrix} p_1 & \dots & p_M \end{bmatrix}^T \quad (20)$$

$$\boldsymbol{\sigma} = \begin{bmatrix} \sigma_1 & \dots & \sigma_N \end{bmatrix}^T \quad (21)$$

$$\mathbf{p} = \begin{bmatrix} \tilde{\mathbf{p}}^T & \boldsymbol{\sigma}^T \end{bmatrix}^T \quad (22)$$

$$\mathbf{P} = \text{diag}(\mathbf{p}) \quad (23)$$

Alternatively, one may consider the more general $\{r, q\}$ -SPICE formulation¹ [33], [34]

$$\underset{\mathbf{p} \geq 0}{\text{minimize}} \quad \mathbf{y}^* \mathbf{R}^{-1} \mathbf{y} + \|\mathbf{p}\|_r + \|\boldsymbol{\sigma}\|_q \quad (24)$$

Using the wideband dictionary over \mathbf{B} in (17) or (24) will allow for much smaller dictionaries as opposed to using ordinary sinusoidal dictionaries. Many other sparse reconstruction techniques may be extended similarly. Generally, the wideband dictionary may be used either as an energy detector which finds the parts of the spectrum that have most energy, or in a zooming procedure similar to the one described above.

V. NUMERICAL EXAMPLES

In this section, we proceed to examine the performance of the proposed method, initially illustrating that the use of a two-stage wideband estimator will have the same estimation quality as when using the ordinary (one-stage) narrowband LASSO estimator.

¹In this formulation, we assume that the columns of the dictionaries are normalized to have unit norm.

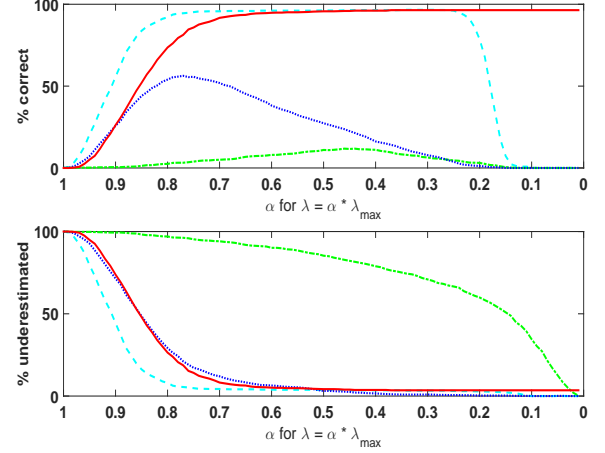


Fig. 7: The probability of (top) correctly estimating and (bottom) underestimating the number of spectral lines, for the (single-stage) narrowband dictionary, using $P = 1000$ elements (cyan, dashed) and $P = 75$ elements (green, dot-dashed), and for the initial wideband dictionary, using $B_1 = 75$ elements (blue, dotted), and the (two-stage) wideband dictionary, using $B_1 = 75$ elements, together with $B_2 = 25$ elements per activated bands in the refining dictionary (red, solid).

A. One-dimensional data

We initially considered a signal consisting of $N = 75$ samples containing $K = 3$ (complex-valued) sinusoids corrupted by a zero-mean white Gaussian noise with signal-to-noise ratio (SNR) of SNR= 10dB. In each simulation, the sinusoidal frequencies are drawn from a uniform distribution, over $[0, 1)$, with all amplitudes having unit magnitude and phases drawn from a uniform distribution over $[0, 2\pi)$. The performance is then computed using three different dictionaries, namely the (ordinary) narrowband dictionary, \mathbf{D} , with $P = 1000$ and $P = 75$ elements, respectively, and the proposed wideband dictionary, \mathbf{B} , using $B_1 = 75$ elements, followed by a second-stage narrowband dictionary using $B_2 = 25$ elements per active band. For each dictionary, we evaluate the performance for varying values of the user parameter α using $\lambda = \alpha \lambda_{\max}$, where $\lambda_{\max} = \max_i |\mathbf{x}_i^H \mathbf{y}_i|$ is the smallest tuning parameter value for which all coefficients in the solution are zero [19]. Here, \mathbf{x}_i denotes either the i th column of the \mathbf{D} dictionary or the i th column of the \mathbf{B} dictionary. Each estimated result is then compared to the ground truth, counting the number

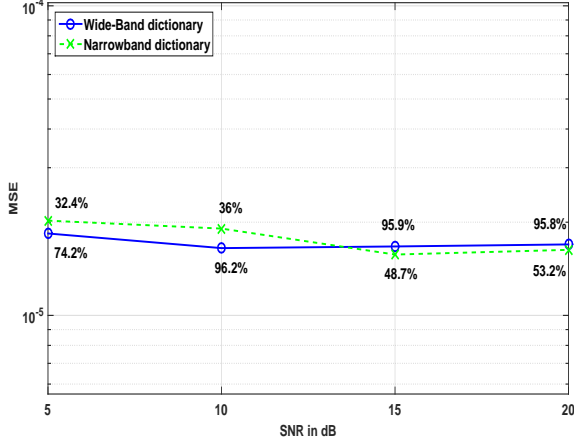


Fig. 8: Mean-square error curves for different SNR levels for the single-stage narrowband dictionary, using $P = 100$, as compared to the two-stage dictionary, using $B_1 = 20$ integrated wideband elements in the first stage, followed by $B_2 = 5$ wideband elements in the second stage. The percentage of correct model order estimation (excluding outliers) is shown as a percentage on top of the corresponding MSE value.

of correctly estimated and underestimated model orders. The results are shown in Figure 7. As can be seen from the figure, the best results are achieved when $\alpha \leq 0.65$, in which case the proposed wideband dictionary, using $B_1 = 75$ bands, followed by a second stage narrowband dictionary, with $B_2 = 25$ for each activated band, have similar performance to the narrowband dictionary using $P = 1000$ dictionary elements.

Proceeding, we assess the mean-square error (MSE) for the two-stage dictionary, showing the MSE as a function of SNR for the first-stage wideband dictionary, B_1 , and second-stage wideband refining dictionary, B_2 . Here, and in the following, we consider situations where the number of elements in the dictionary is less than number of samples. As was described before, this is a situation where the performance of narrowband dictionaries can deteriorate seriously. For this experiment, we considered a signal with $N = 300$ samples containing $K = 2$ (complex-valued) sinusoids, being corrupted by different levels of zero-mean white Gaussian noise with SNR in the range $[5, 20]$ dB. Figure 8 shows the resulting MSE for the LASSO estimator for the estimates with correctly estimated

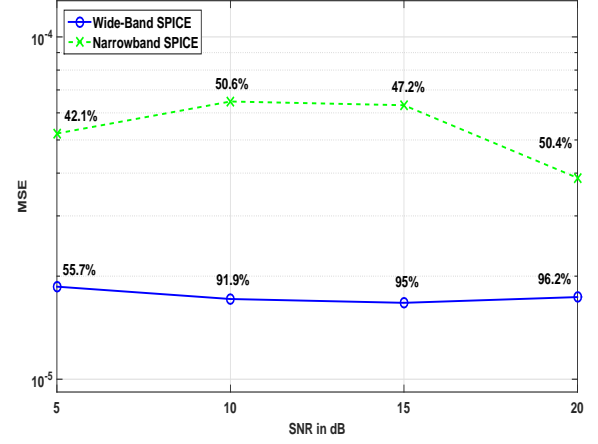


Fig. 9: Mean-square error curves for different SNR levels for the single-stage narrowband dictionary, using $P = 100$, as compared to the two-stage dictionary, using $B_1 = 20$ integrated wideband elements in the first stage, followed by $B_2 = 5$ wideband elements in the second stage. The percentage of correct model order estimation (excluding outliers) is shown as a percentage on top of the corresponding MSE value.

model order; for runs with the correct model order estimation we also removed outliers from the final MSE calculation. We consider an estimate as an outlier if $|f - \hat{f}| > \Delta f$, where Δf was defined as two times the possible resolution, where possible resolution is defined as $1/P$ for the narrowband dictionary and $1/(B_1 \cdot B_2)$ for the wideband dictionary. Figure 9 shows the MSE for the same experiment done using the SPICE estimator. The number of outliers removed for the LASSO estimator was: 4, 0, 0, 0 for the wideband dictionary and 7, 16, 10 and 11 for the narrowband dictionary (corresponding to SNRs of 5, 10, 15, and 20 dB). The number of outliers removed for the SPICE estimator was; 17, 1, 1, 0 for the wideband dictionary and 52, 80, 117, and 103 for the narrowband dictionary. As can be seen from the figures, the two-stage dictionary using a wideband dictionary using $B_1 = 20$ bands, followed by a refining dictionary using $B_2 = 5$ wideband elements, achieves the same performance as the single-stage narrowband dictionary using $P = 100$ elements in terms of resolution. However, the narrow-band dictionary will for this case fail to reliably restore the signal with reconstruction success rates of merely 30 – 50%.

Settings	Complexity ratio
$D = 1000, N = 200, K = 3$	1
$B_1 = 20, B_2 = 5$	897
$B_1 = 20, B_2 = 40$	27
$D = 1000, N = 400, K = 3$	26
$B_1 = 10, B_2 = 10, B_3 = 5$	1000

TABLE I: Complexity reduction compared to using the full dictionary and the distance between the final grid for different settings. Here, D indicates the narrowband dictionary, whereas B_1, B_2 indicates the two-stage dictionary using B_1 wideband elements in the first stage, followed by B_2 wideband elements in the second-stage.

Table I shows the corresponding complexity cost of some of the different settings in the numerical section. Note that to simplify the comparison this is the complexity of solving the ADMM without utilizing any structures of the dictionary matrices.

Next, we consider non-uniformly sampled data with $N = 400$ samples, for $K = 2$ sinusoids. For this experiment, we also added a third estimation step for the iterative wideband dictionary. After initial estimation with $B_1 = 10$ wideband dictionary elements, we zoom into the active bands with $B_2 = 10$ dictionary elements per active band, and then once again with $B_3 = 5$ dictionary elements. In spite of the three stage zooming, the method requires considerably less computational operations as compared to using a corresponding narrowband dictionary, but results in better performance both in terms of resolution and model-order accuracy. The resulting MSEs are shown in Figure 10. All results are computed using 1000 Monte-Carlo simulations.

B. Two-dimensional data

In this subsection, we present results on a 2-D data set. In this example, each dimension is sampled uniformly with $N = 100$ samples. We compare a narrowband dictionary with $P = 49$ elements per dimension with the wideband dictionary using $B_1 = 7$ bands per dimension in the first step and a wideband dictionary with $B_2 = 7$ elements per active band in a second (zooming) step. Here, we use two separate wideband dictionaries, the first, \mathbf{B} , using integrated dictionary elements as defined in (7), and the second, \mathbf{B}_{DPSS} , which contains elements based on DPSS. For the DPSS-based dictionary, we used a sequence length of $Q = 100$ and $W = 1/2.1$. Using $W < 1/2.1$ results in dictionary elements

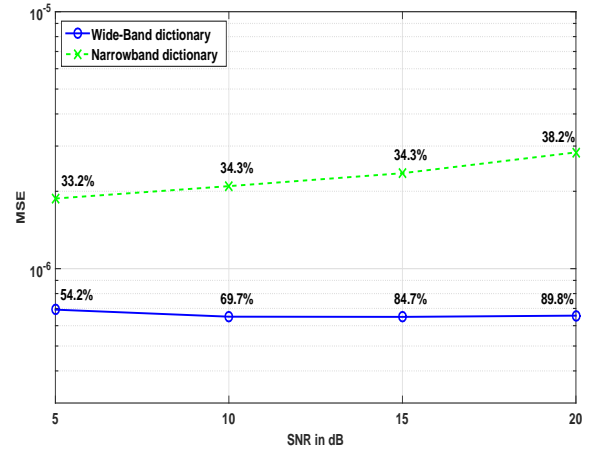


Fig. 10: Signal estimation for non-uniform sampling: mean-square error curves for different SNR levels for the single-stage narrowband dictionary, using $P = 200$ elements, as compared to the three-stage dictionary, using $B_1 = 10$ integrated wideband elements in the first stage, followed by $B_2 = 10$ and $B_3 = 5$ wideband dictionaries in the second stage and third stage per active band detected in the previous stage. The correct model order estimations are shown in percentage above each point.

which concentrate energy in a more narrow band and are therefore not suitable for the dictionary with $B_1 = B_2 = 7$ elements. We considered a signal containing $K = 2$ (complex-valued) sinusoids per dimension, with the signal being corrupted by a zero-mean white Gaussian noise. In each simulation, the sinusoidal frequencies are drawn from a uniform distribution, over $[0, 1)$, with all the amplitudes having unit magnitude. The two dictionaries are compared against each other based on the MSE performance in the same manner as in the previous subsection, with the MSE being calculated as the average value for both dimensions if the model order estimate for the iteration was correct. Outliers are removed before the MSE calculation. The number of outliers removed was: 6, 15, 25, 30 for the \mathbf{B}_{DPSS} dictionary and 20, 22, 17 and 24 for the narrowband dictionary (corresponding to SNRs of 5, 10, 15, and 20 dB). The wideband dictionary \mathbf{B} did not result in any outliers. The percentages of correct model order estimates are shown for each SNR value. Figure 11 shows the resulting MSE curves. It can be seen that the wideband dictionary with integrated sinusoids outperforms the DPSS-

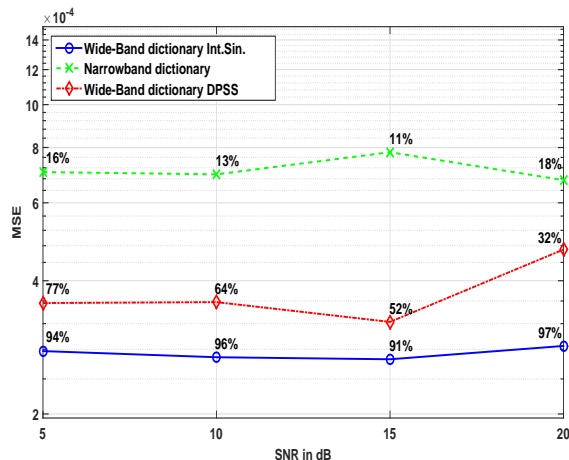


Fig. 11: Signal estimation in two dimensions: mean-square error curves for different SNR levels for the single-stage narrowband dictionary \mathbf{D} , using $P = 100$ per dimension, as compared to the two-stage dictionaries (DPSS based and integrated sinusoids based), using $B_1 = 7$ wideband elements in the first stage, followed by $B_2 = 7$ wideband elements in the second stage (per active band).

based wideband dictionary both in terms of MSE and model-order accuracy. Comparing to using the narrowband dictionary, it can be seen that both wideband dictionaries outperform it both in terms of MSE and model-order estimation. Also in this example, the wideband dictionaries provide a considerable reduction in computational complexity as well as a robustness in terms of estimating off-grid components. All results are computed using 100 Monte-Carlo simulations.

Using the same setup as described above we also evaluated the performance of the proposed approach when the number of sinusoids to detect is higher. Again, we considered the ordinary narrowband dictionary, \mathbf{D} , and the wideband dictionary, \mathbf{B} , from the previous experiment. We calculated the percentage of correct model order estimation for signals with $K = 4, 6, 8$, and 10 (complex-valued) sinusoids. The results were computed using 100 Monte-Carlo simulations; the correct model order estimation percentages for different SNR levels are shown in Figure 12. The best regularization parameters λ for solving the LASSO for each case were found beforehand with the grid-search method. For this, we selected the range of parameter $\alpha \in [0.7, 0.05]$ with the step-size 0.05 and ran 100 Monte-Carlo

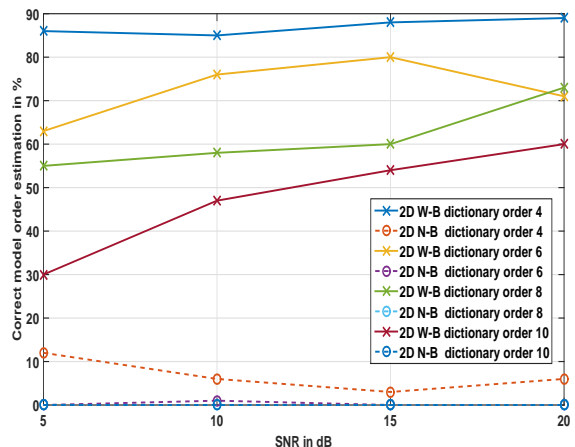


Fig. 12: Percentage of correct model order estimations for different number of sinusoids and for different SNR levels for wideband dictionary (W-B) and narrowband dictionary (N-B).

simulations for each model order and then picked the best parameter for the selected model order based on model order accuracy. For the two-step wideband dictionary, a grid-search was done for the set of α parameter for the both stages. It can be clearly seen that for situations where the number of elements in the dictionary is lower than the number of samples, the narrow-band dictionary fails to produce any meaningful results.

VI. CONCLUSION

In this paper, we have introduced a wideband dictionary framework, allowing for a computationally efficient reconstruction of sparse signals. Wideband dictionary elements are formed as spanning bands of the considered parameter space. In the first stage, one may typically use a coarse grid using the integrated wideband dictionary locating the bands of interest, whereafter non-active parts of the parameter space are discarded. In the next stage, a refining dictionary can be used to more precisely determine the parameters of interest on the active bands from the previous step, allowing for an iterative zooming procedure. The technique is illustrated for the problem of estimating multi-dimensional sinusoids corrupted by Gaussian noise, showing that the same accuracy can be achieved, although at a computationally substantially lower cost and with much less risk of missing any off-grid components. The proposed framework is here

illustrated for the LASSO and SPICE estimators, but other sparse reconstruction techniques may be extended similarly.

REFERENCES

- [1] M. Unser and P. Tafti, *An introduction to sparse stochastic processes*, Cambridge University Press, 2013.
- [2] M. Elad, *Sparse and Redundant Representations*, Springer, 2010.
- [3] E. J. Candès and M. B. Wakin, “An Introduction To Compressive Sampling,” *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, March 2008.
- [4] E. J. Candès, J. Romberg, and T. Tao, “Robust Uncertainty Principles: Exact Signal Reconstruction From Highly Incomplete Frequency Information,” *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.
- [5] D.L. Donoho, “Compressed Sensing,” *IEEE Trans. Inf. Theory*, vol. 52, pp. 1289–1306, 2006.
- [6] M. A. Herman and T. Strohmer, “Genral Deviants: An Analysis of Perturbations in Compressed Sensing,” *IEEE J. Sel. Topics in Signal Processing*, vol. 4, no. 2, pp. 342–349, April 2010.
- [7] Y. Chi, L. L. Scharf, A. Pezeshki, and A. R. Calderbank, “Sensitivity to Basis Mismatch in Compressed Sensing,” *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 2182–2195, May 2011.
- [8] P. Stoica and P. Babu, “Sparse Estimation of Spectral Lines: Grid Selection Problems and Their Solutions,” *IEEE Trans. Signal Process.*, vol. 60, no. 2, pp. 962–967, Feb. 2012.
- [9] G. Tang, B. N. Bhaskar, P. Shah, and B. Recht, “Compressed Sensing Off the Grid,” *IEEE Trans. Inform. Theory*, vol. 59, no. 11, pp. 7465–7490, Nov 2013.
- [10] Y. Chi and Y. Chen, “Compressive Two-Dimensional Harmonic Retrieval via Atomic Norm Minimization,” *IEEE Trans. Signal Process.*, vol. 63, no. 4, pp. 1030–1042, Feb 2015.
- [11] Z. Yang and L. Xie, “Enhancing Sparsity and Resolution via Reweighted Atomic Norm Minimization,” *IEEE Trans. Signal Process.*, vol. 64, no. 4, pp. 995–1006, Feb 2016.
- [12] V. Chandrasekaran, B. Recht, P. A. Parrilo, and A. S. Willsky, “The Convex Geometry of Linear Inverse Problems,” *Foundations of Computational Mathematics*, vol. 12, no. 6, pp. 805–849, Dec 2012.
- [13] J. J. Fuchs, “On the Use of Sparse Representations in the Identification of Line Spectra,” in *17th World Congress IFAC*, Seoul, Jul 2008, pp. 10225–10229.
- [14] P. Stoica, P. Babu, and J. Li, “New method of sparse parameter estimation in separable models and its use for spectral analysis of irregularly sampled data,” *IEEE Trans. Signal Process.*, vol. 59, no. 1, pp. 35–47, Jan 2011.
- [15] P. Stoica and P. Babu, “SPICE and LIKES: Two hyperparameter-free methods for sparse-parameter estimation,” *Signal Processing*, vol. 92, no. 7, pp. 1580–1590, July 2012.
- [16] I. F. Gorodnitsky and B. D. Rao, “Sparse Signal Reconstruction from Limited Data Using FOCUSS: A Re-weighted Minimum Norm Algorithm,” *IEEE Trans. Signal Process.*, vol. 45, no. 3, pp. 600–616, March 1997.
- [17] S. I. Adalbjörnsson, A. Jakobsson, and M. G. Christensen, “Multi-Pitch Estimation Exploiting Block Sparsity,” *Elsevier Signal Processing*, vol. 109, pp. 236–247, April 2015.
- [18] L. E. Ghaoui, V. Viallon, and T. Rabbani, “Safe Feature Elimination for the LASSO and Sparse Supervised Learning Problems,” 2011, Publication: eprint arXiv:1009.4219v2.
- [19] R. Tibshirani, J. Bienand, J. Friedman, T. Hastieand N. Simon, J. Taylor, and R. J. Tibshirani, “Strong rules for discarding predictors in lasso-type problems,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 74, no. 2, pp. 245–266, 2012.
- [20] Z. J. Xiang, Y. Wang, and P. J. Ramadge, “Screening Tests for Lasso Problems,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PP, no. 99, 2016.
- [21] A. Bonnefoy, V. Emiya, L. Ralaivola, and R. Gribonval, “A Dynamic Screening Principle for the Lasso,” in *Proceedings of the 22nd European Signal Processing Conference*, Lisbon, Portugal, 1-5 September 2014.
- [22] O. Fercoq, A. Gramfort, and J. Salmon, “Mind the Duality Gap: Safe Rules for the Lasso,” 2015, Publication: eprint arXiv:1505.03410v3.
- [23] J. Liu, Z. Zhao, J. Wang, and J. Ye, “Safe Screening With Variational Inequalities and Its Application to LASSO,” 2014, Publication: eprint arXiv:1307.7577v3.
- [24] Z. Yang and L. Xie, “Frequency-Selective Vandermonde Decomposition of Toeplitz Matrices With Applications,” 2016, Publication: eprint arXiv:1605.02431.
- [25] R. Tibshirani, “Regression shrinkage and selection via the Lasso,” *Journal of the Royal Statistical Society B*, vol. 58, no. 1, pp. 267–288, 1996.
- [26] P. Stoica, P. Babu, and J. Li, “SPICE : a novel covariance-based sparse estimation method for array processing,” *IEEE Trans. Signal Process.*, vol. 59, no. 2, pp. 629–638, Feb. 2011.
- [27] P. Stoica, D. Zachariah, and L. Li, “Weighted SPICE: A Unified Approach for Hyperparameter-Free Sparse Estimation,” *Digit. Signal Process.*, vol. 33, pp. 1–12, October 2014.
- [28] S. Sahnoun, E. H. Djermoune, and D. Brie, “Sparse Modal Estimation of 2-D NMR Signals,” in *38th IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, Vancouver, Canada, May 26-31 2013.
- [29] J. Swärd, S. I. Adalbjörnsson, and A. Jakobsson, “High Resolution Sparse Estimation of Exponentially Decaying N-dimensional Signals,” *Elsevier Signal Processing*, vol. 128, pp. 309–317, Nov 2016.
- [30] D. Slepian, “Prolate Spheroidal Wave Functions, Fourier Analysis, and Uncertainty V: the Discrete Case,” *The Bell System Technical Journal*, vol. 57, no. 5, pp. 1371–1430, May-June 1978.
- [31] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, “Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers,” *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, Jan. 2011.
- [32] G. H. Golub and C. F. Van Loan, *Matrix Computations*, The John Hopkins University Press, 4th edition, 2013.
- [33] J. Swärd, S. I. Adalbjörnsson, and A. Jakobsson, “A Generalization of the Sparse Iterative Covariance-based Estimator,” in *42nd IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, New Orleans, USA, March, 5-9 2017.
- [34] J. Swärd, S. I. Adalbjörnsson, and A. Jakobsson, “Generalized Sparse Covariance-based Estimation,” *Elsevier Signal Processing*, 2017, Accepted for publication.